日本国特許庁
JAPAN PATENT OFFICE

13.12.02

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出願年月日
Date of Application: ２００２年　１月１６日

出願番号
Application Number: 特願２００２－００７２８３

[ST.10/C]: [ＪＰ２００２－００７２８３]

出願人
Applicant(s): シャープ株式会社

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

２００３年　１月２８日

特許庁長官
Commissioner,
Japan Patent Office 太田信一郎

出証番号　出証特２００２－３１０７９８０

【書類名】　特許願
【整理番号】　181093
【提出日】　平成14年 1月16日
【あて先】　特許庁長官殿
【国際特許分類】　G10L 3/00
G10L 5/06
【発明者】
【住所又は居所】　大阪府大阪市阿倍野区長池町２２番２２号　シャープ株式会社内
【氏名】　鶴田　彰
【特許出願人】
【識別番号】　000005049
【住所又は居所】　大阪府大阪市阿倍野区長池町２２番２２号
【氏名又は名称】　シャープ株式会社
【代理人】
【識別番号】　100062144
【弁理士】
【氏名又は名称】　青山　葆
【選任した代理人】
【識別番号】　100086405
【弁理士】
【氏名又は名称】　河宮　治
【選任した代理人】
【識別番号】　100084146
【弁理士】
【氏名又は名称】　山崎　宏
【手数料の表示】
【予納台帳番号】　013262
【納付金額】　21,000円

【提出物件の目録】

【物件名】　明細書　1

【物件名】　図面　1

【物件名】　要約書　1

【包括委任状番号】　0003090

【プルーフの要否】　要

【書類名】　明細書

【発明の名称】　連続音声認識装置および連続音声認識方法、連続音声認識プログラム、並びに、プログラム記録媒体

【特許請求の範囲】

【請求項１】　隣接するサブワードに依存して決定されるサブワードを認識単位とすると共に、サブワード環境に依存する環境依存音響モデルを用いて、連続的に発声された入力音声を認識する連続音声認識装置であって、

上記入力音声を分析して特徴パラメータの時系列を得る音響分析部と、

語彙中の各単語が、サブワードのネットワークあるいはサブワードの木構造として格納された単語辞書と、

単語間の接続情報を表す言語モデルが格納された言語モデル格納部と、

上記環境依存音響モデルが、当該環境依存音響モデルの状態系列のうち、複数のサブワードモデルの状態系列をまとめて木構造化して成るサブワード状態木として格納されている環境依存音響モデル格納部と、

上記環境依存音響モデルであるサブワード状態木,単語辞書および言語モデルを参照して上記サブワードの仮説を展開すると共に、上記特徴パラメータの時系列と上記展開された仮説との照合を行い、単語の終端に該当する仮説に関する単語,累積スコアおよび始端開始フレームを含む単語情報を単語ラティスとして出力する照合部と、

上記単語ラティスに対する探索を行って認識結果を生成する探索部

を備えたことを特徴とする連続音声認識装置。

【請求項２】　請求項１に記載の連続音声認識装置において、

上記環境依存音響モデル格納部に格納されている環境依存音響モデルは、中心サブワードが前後のサブワードに依存する環境依存音響モデルのうち、先行サブワードおよび中心サブワードが同じサブワードモデルの状態系列を木構造化したサブワード状態木であることを特徴とする連続音声認識装置。

【請求項３】　請求項２に記載の連続音声認識装置において、

上記環境依存音響モデルは、複数のサブワードモデルで状態を共有している状態共有モデルであることを特徴とする連続音声認識装置。

【請求項４】　請求項１に記載の連続音声認識装置において、

上記照合部は、上記サブワード状態木を参照して仮説を展開する際に、上記単語辞書および言語モデルから得られる接続可能なサブワード情報を用いて、上記仮説であるサブワード状態木を構成する状態のうち、互いに接続可能な状態にフラグを付すようになっていることを特徴とする連続音声認識装置。

【請求項５】　請求項１に記載の連続音声認識装置において、

上記照合部は、上記照合を行う際に、上記特徴パラメータの時系列に基づいて上記展開された仮説のスコアを算出すると共に、このスコアの閾値あるいは仮説数を含む基準に従って上記仮説の枝刈りを行うようになっていることを特徴とする連続音声認識装置。

【請求項６】　隣接するサブワードに依存して決定されるサブワードを認識単位とすると共に、サブワード環境に依存する環境依存音響モデルを用いて、連続的に発声された入力音声を認識する連続音声認識方法であって、

音響分析部によって、上記入力音声を分析して特徴パラメータの時系列を得、

照合部によって、上記環境依存音響モデルの状態系列を木構造化して成るサブワード状態木、語彙中の各単語がサブワードのネットワークあるいはサブワードの木構造として記述された上記単語辞書、および、単語間の接続情報を表す言語モデルを参照して、上記サブワードの仮説を展開すると共に、上記特徴パラメータの時系列と上記展開された仮説との照合を行って、単語の終端に該当する仮説に関する単語,累積スコアおよび始端開始フレームを含む単語情報を単語ラティスとして生成し、

探索部によって、上記単語ラティスに対する探索を行って認識結果を生成することを特徴とする連続音声認識方法。

【請求項７】　コンピュータを、請求項１に記載の音響分析部,単語辞書,言語モデル格納部,環境依存音響モデル格納部,照合部および探索部として機能させることを特徴とする連続音声認識プログラム。

【請求項８】　請求項７に記載の連続音声認識プログラムが記録されたことを特徴とするコンピュータ読出し可能なプログラム記録媒体。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】

この発明は、音素環境依存音響モデルを用いて高精度に認識を行う連続音声認識装置および連続音声認識方法、連続音声認識プログラム、並びに、連続音声認識プログラムを記録したプログラム記録媒体に関する。

【０００２】

【従来の技術】

一般に、大語彙連続音声認識で用いる認識単位としては、認識対象語彙の変更や大語彙への拡張が容易であることから、音節や音素等の単語より小さいサブワードと呼ばれる認識単位が用いられることが多い。さらに、調音結合等の影響を考慮するためには、前後の環境(コンテキスト)に依存したモデルが有効であることが知られている。例えば、前後一つずつの音素に依存したトライフォンモデルと呼ばれる音素モデルが広く使用されている。

【０００３】

また、連続的に発声された音声を認識する連続音声認識方法の一つとして、語彙中の各単語をサブワードのネットワークや木構造等で記述したサブワード表記辞書と、単語の接続の制約を記述した文法または統計的言語モデルの情報とに従って、単語を連結して認識結果を得る方法がある。

【０００４】

これらのサブワードを認識単位とした連続音声認識技術については、例えば、刊行物「音声認識の基礎(下)」古井貞熙監訳に詳しく説明されている。

【０００５】

上述したごとく、環境に依存したサブワードを用いて連続音声認識を行う場合には、単語内だけではなく単語間においても音素環境依存型の音響モデルを用いた方が、認識精度がよいことが知られている。しかしながら、単語の始終端に用いる音響モデルは前後に接続する単語に依存するため、音素環境に依存しない音響モデルを用いる場合に比べて、処理が複雑になると共に処理量が大幅に増えてしまう。

【０００６】

以下、単語辞書と言語モデルと音素環境依存音響モデルを参照して、単語履歴毎に木を動的に生成する方法について、具体的に説明する。

【０００７】

例えば、「朝の天気…」という発声に対して、「朝(a;s;a)」という単語の最後の音素/ａ/を考える場合、図３に示す単語辞書の情報から得られる単語「朝日(a;s;a;h;i)」における３番目の音素/ａ/とその前後に続く音素とから成るトライフォン“s;a;h”と、図４に示す言語モデルの情報から得られる単語「の(n;o)」とその前に続く単語「朝(a;s;a)」との連鎖「朝の(a;s;a;n;o)」における３番目の音素/ａ/とその前後に続く音素とから成るトライフォン“s;a;n”とについて、仮説を展開する必要がある。この例の場合は２つの仮説を展開するだけでよいが、より複雑な文法や統計的言語モデルを用いる場合には、単語の終端で多くの単語につながる可能性がある。そして、その場合には、それらの先頭の音素に依存して、例えば図２(b)に示すような先行音素と中心音素と後続音素からなるトライフォンの状態系列を用いて、図５(b)に示すように多くの仮説を展開する必要がある。

【０００８】

この問題に対し、単語内には音素環境依存の音響モデルを用いる一方、単語境界では環境に依存しない音響モデルを使用する連続音声認識方式が、特開平５－２２４６９２号公報に開示されている。この連続音声認識方式によれば、単語間での処理量の増大を抑えることができる。また、認識対象語彙中の各単語について、前後の単語に依存せずに決まる音響モデル系列を認識単語として記述した認識単語辞書と、単語境界において前後の単語に依存して記述した単語間単語辞書とを用いて照合する連続音声認識方式が、特開平１１－４５０９７号公報に開示されている。この連続音声認識方式によれば、単語境界に音素環境依存の音響モデルを用いても処理量の増大を抑えることができるのである。

【０００９】

【発明が解決しようとする課題】

しかしながら、上記従来の連続音声認識方式においては、以下のような問題がある。すなわち、特開平５－２２４６９２号公報に開示された連続音声認識方式においては、単語内には音素環境依存の音響モデルを用い、単語境界では環境に

依存しない音響モデルを用いている。したがって、単語境界での処理量の増大を抑えることはができるが、その一方において、単語境界に用いる音響モデルの精度が低いために、特に大語彙の連続音声認識の場合には認識性能の低下を招く恐れがある。

【００１０】

これに対して、特開平１１－４５０９７号公報に開示された連続音声認識方式においては、前後の単語に依存せずに決まる音響モデル系列を認識単語として記述した認識単語辞書と、単語境界において前後の単語に依存して記述した単語間単語辞書を用いて照合を行うようにしている。したがって、単語境界にも音素環境依存の音響モデルを用いることによって精度を確保しながら、大語彙の場合でも単語境界での処理量の増大を抑えることができるのである。しかしながら、一般に、単語のスコアや境界はそれ以前の単語の影響を受けるので、複数の認識単語が単語間単語を共有すると、図９(a)に示すように認識単語“k;o;k”及び“s;o;k”と単語間単語“o”との境界の履歴が考慮されないので、図９(b)に示すように単語の境界履歴を考慮した場合に比して、性能の低下を招く恐れがある。また、例えば助詞の“を(/o/と発声)”等のように、認識単語辞書と単語間単語辞書とに分割することができない単語についは開示されてはいない。

【００１１】

そこで、この発明の目的は、単語境界にも音素環境依存音響モデルを用いて精度を確保しつつ、大語彙の連続音声認識時にも単語境界での処理量の増大を抑えることができる連続音声認識装置および連続音声認識方法、連続音声認識プログラム、並びに、連続音声認識プログラムを記録したプログラム記録媒体を提供することにある。

【００１２】

【課題を解決するための手段】

上記目的を達成するため、第１の発明は、隣接するサブワードに依存して決定されるサブワードを認識単位とすると共に,サブワード環境に依存する環境依存音響モデルを用いて,連続的に発声された入力音声を認識する連続音声認識装置であって、入力音声を分析して特徴パラメータの時系列を得る音響分析部と、語

彙中の各単語が,サブワードのネットワークあるいはサブワードの木構造として格納された単語辞書と、単語間の接続情報を表す言語モデルが格納された言語モデル格納部と、上記環境依存音響モデルが,当該環境依存音響モデルの状態系列のうち,複数のサブワードモデルの状態系列をまとめて木構造化して成るサブワード状態木として格納されている環境依存音響モデル格納部と、上記環境依存音響モデルであるサブワード状態木,上記単語辞書および言語モデルを参照して上記サブワードの仮説を展開すると共に,上記特徴パラメータの時系列と上記展開された仮説との照合を行い,単語の終端に該当する仮説に関する単語,累積スコア及び始端開始フレームを含む単語情報を単語ラティスとして出力する照合部と、上記単語ラティスに対する探索を行って認識結果を生成する探索部を備えたことを特徴としている。

【００１３】

上記構成によれば、サブワード環境に依存する環境依存音響モデルを木構造化したサブワード状態木,単語辞書および言語モデルを参照して、サブワードの仮説を展開するようにしている。したがって、次に続く単語の先頭サブワードに関係無く１つの仮説を展開すればよく、全仮説における状態の総数を削減することができる。すなわち、仮説の展開処理量を大幅に削減でき、単語内および単語境界に関係なく、仮説の展開が容易になるのである。さらに、照合部によって、上記音響分析部からの特徴パラメータ系列と上記展開された仮説との照合を行う際における照合処理量が大幅に削減される。

【００１４】

また、１実施例では、上記第１の発明の連続音声認識装置において、上記環境依存音響モデル格納部に格納されている環境依存音響モデルは、中心サブワードが前後のサブワードに依存する環境依存音響モデルのうち、先行サブワードおよび中心サブワードが同じサブワードモデルの状態系列を木構造化したサブワード状態木である。

【００１５】

この実施例によれば、先行サブワードおよび中心サブワードが同じサブワードモデルの状態系列を木構造化したサブワード状態木を用いて、上記仮説を展開し

ている。したがって、次の仮説を展開する場合には、終端仮説における中心サブワードのみに注目して対応する先行サブワードを有するサブワード状態木を展開すればよい。つまり、後続サブワードが複数あってもより少ない仮説を展開すればよく、仮説の展開が容易である。

【００１６】

また、１実施例では、上記第１の発明の連続音声認識装置において、上記環境依存音響モデルは、複数のサブワードモデルで状態を共有している状態共有モデルである。

【００１７】

この実施例によれば、複数のサブワードモデルによって状態を共有することによって、木構造化した際に共有している状態を一つにまとめることができ、ノード数を削減することができる。したがって、上記照合部による照合時における処理量が大幅に削減される。

【００１８】

また、１実施例では、上記第１の発明の連続音声認識装置において、上記照合部は、上記サブワード状態木を参照して仮説を展開する際に、上記単語辞書および言語モデルから得られる接続可能なサブワード情報を用いて、上記仮説であるサブワード状態木を構成する状態のうち、互いに接続可能な状態にフラグを付すようになっている。

【００１９】

この実施例によれば、上記展開された仮説を構成するサブワード状態木の状態のうち、互いに接続可能な状態のみにフラグを付けるようにしたので、上記照合の際にビタビ計算を行う必要がある状態が限定されて、照合処理量が更に簡単になる。

【００２０】

また、１実施例では、上記第１の発明の連続音声認識装置において、上記照合部は、上記照合を行う際に、上記特徴パラメータの時系列に基づいて上記展開された仮説のスコアを算出すると共に、このスコアの閾値あるいは仮説数を含む基準に従って上記仮説の枝刈りを行うようになっている。

【００２１】

この実施例によれば、上記照合時に仮説の枝刈りを行うので、単語となる可能性が低い仮説が削除されて、以後の照合処理量が大幅に削減される。

【００２２】

また、第２の発明は、隣接するサブワードに依存して決定されるサブワードを認識単位とすると共に，サブワード環境に依存する環境依存音響モデルを用いて，連続的に発声された入力音声を認識する連続音声認識方法であって、音響分析部によって，上記入力音声を分析して特徴パラメータの時系列を得、照合部によって，上記環境依存音響モデルの状態系列を木構造化して成るサブワード状態木，語彙中の各単語がサブワードのネットワークあるいはサブワードの木構造として記述された上記単語辞書，および，単語間の接続情報を表す言語モデルを参照して，上記サブワードの仮説を展開すると共に，上記特徴パラメータの時系列と上記展開された仮説との照合を行って，単語の終端に該当する仮説に関する単語，累積スコアおよび始端開始フレームを含む単語情報を単語ラティスとして生成し、探索部によって，上記単語ラティスに対する探索を行って認識結果を生成することを特徴としている。

【００２３】

上記構成によれば、上記第１の発明の場合と同様に、環境依存音響モデルを木構造化したサブワード状態木を参照して仮説を展開するので、次に続く単語の先頭サブワードに関係無く１つの仮説を展開すればよく、単語内および単語境界に関係なく仮説の展開が容易になるのである。さらに、特徴パラメータ系列と上記展開された仮説との照合を行う際における照合処理量が大幅に削減される。

【００２４】

また、第３の発明の連続音声認識プログラムは、コンピュータを、上記第１の発明における音響分析部，単語辞書，言語モデル格納部，環境依存音響モデル格納部，照合部および探索部として機能させることを特徴としている。

【００２５】

上記構成によれば、上記第１の発明の場合と同様に、次に続く単語の先頭サブワードに関係無く１つの仮説を展開すればよく、単語内および単語境界に関係な

く仮説の展開が容易になる。さらに、特徴パラメータ系列と上記展開された仮説との照合を行う際における照合処理量が大幅に削減される。

【００２６】

また、第４の発明のプログラム記録媒体は、上記第３の発明の連続音声認識プログラムが記録されたことを特徴としている。

【００２７】

上記構成によれば、上記第１の発明の場合と同様に、次に続く単語の先頭サブワードに関係無く１つの仮説を展開すればよく、単語内および単語境界に関係なく仮説の展開が容易になる。さらに、特徴パラメータ系列と上記展開された仮説との照合を行う際における照合処理量が大幅に削減される。

【００２８】

【発明の実施の形態】

以下、この発明を図示の実施の形態により詳細に説明する。図１は、本実施の形態の連続音声認識装置におけるブロック図である。この連続音声認識装置は、音響分析部１,前向き照合部２,音素環境依存音響モデル格納部３,単語辞書４,言語モデル格納部５,仮説バッファ６,単語ラティス格納部７および後向き探索部８で構成される。

【００２９】

図１において、入力音声は、音響分析部１によって、特徴パラメータの系列に変換されて前向き照合部２に出力される。前向き照合部２では、音素環境依存音響モデル格納部３に格納された音素環境依存音響モデル,言語モデル格納部５に格納された言語モデルおよび単語辞書４を参照して、仮説バッファ６上に音素仮説を展開する。そして、上記音素環境依存音響モデルを用いて、上記展開された音素仮説と特徴パラメータ系列との照合をフレーム同期ビタビビームサーチによって行い、単語ラティスを生成して単語ラティス格納部７に格納する。

【００３０】

上記音素環境依存音響モデルとしては、トライフォンモデルと呼ばれる前後一つずつの音素環境を考慮した隠れマルコフモデル(HMM)を用いている。すなわち、上記サブワードモデルは音素モデルである。但し、従来においては図２(b)

に示すように中心音素の前後１つずつの先行音素と後続音素とを考慮したトライフォンモデルを３状態の状態系列(状態番号列)で表現していたものを、本実施の形態においては、図２(a)に示すように、先行音素と中心音素とが同じトライフォンモデルの状態系列をまとめて木構造(以下、音素状態木という)化している。図２(b)に示すように、複数のトライフォンモデルで状態を共有している状態共有モデルは、状態系列を木構造化して音素状態木を作成することによって状態数を削減することができ、計算量の削減を行うことができるのである。

【００３１】

上記単語辞書４としては、認識対象語彙の各単語について、その単語の読みを音素系列で表記し、図３に示すように、上記音素系列を木構造化したものを用いる。言語モデル格納部５には、例えば、図４に示すように、文法によって設定された単語間の接続情報が言語モデルとして格納されている。尚、本実施の形態においては、単語の読みを表わす音素系列を木構造化したものを単語辞書４としているが、ネットワーク化したものでも差し支えない。また、言語モデルとして文法モデルを用いたが、統計的言語モデルを用いても差し支えない。

【００３２】

上記仮説バッファ６上には、上述したように、上記前向き照合部２によって、音素環境依存音響モデル格納部３,単語辞書４および言語モデル格納部５が参照されて、図５(a)に示すような音素仮説が順次展開される。後向き探索部８は、言語モデル格納部５に格納された言語モデルおよび単語辞書４を参照しながら、単語ラティス格納部７に格納されている単語ラティスを、例えばＡ＊アルゴリズムを用いて探索することによって、入力音声に対する認識結果を得るようになっている。

【００３３】

以下、上記前向き照合部２によって、上記音素環境依存音響モデル格納部３，単語辞書４および言語モデル格納部５を参照して、仮説バッファ６上に仮説を展開して単語ラティスを生成する方法について、図６に示す前向き照合処理動作フローチャートに従って説明する。

【００３４】

ステップＳ1で、先ず照合を始める前に仮説バッファ６の初期化を行う。そして、無音から各単語の始端に続く“-;-;＊”なる音素状態木が初期仮説として仮説バッファ６にセットされる。ステップＳ2で、上記音素環境依存音響モデルが用いられて、処理対象のフレームにおける特徴パラメータと仮説バッファ６内にある図７(a)に示すような音素仮説との照合が行われ、各音素仮説のスコアが計算される。ステップＳ3で、図７(b)に示すように、上記スコアの閾値あるいは仮説数等に基づいて、仮説１及び仮説４のように音素仮説の枝刈りが行われる。こうして、音素仮説の不必要な増大が防止される。ステップＳ4で、仮説バッファ６内に残っている音素仮説のうち単語終端がアクティブなものについて、単語，累積スコアおよび始端開始フレーム等の単語情報が単語ラティス格納部７に保存される。こうして、単語ラティスが生成されて保存される。ステップＳ5で、図７(b)に示される仮説５および仮説６のように、音素環境依存音響モデル格納部３,単語辞書４および言語モデル格納部５の情報が参照されて、仮説バッファ６内に残っている音素仮説が伸ばされる。ステップＳ6で、当該処理対象フレームは最終フレームであるか否かが判別される。その結果、最終フレームである場合には前向き照合処理動作を終了する。一方、最終フレームでない場合には上記ステップＳ2に戻って、次のフレームの処理に移行する。そして、以後、上記ステップＳ2～ステップＳ6までが繰り返され、上記ステップＳ6において最終フレームであると判別されると前向き照合処理動作を終了する。

【００３５】

以下、上記前向き照合処理動作の際に、先行音素および中心音素が同じであるトライフォンモデルの状態系列が木構造化された音素状態木を用いる場合の効果について説明する。

【００３６】

例えば、「朝の天気…」という発声に対して、「朝(a;s;a)」という単語の最後の音素/ａ/を考える場合に、図３に示す単語辞書４の情報から得られた単語「朝日(a;s;a;h;i)」における３番目の音素/ａ/とその前後に続く音素とから成るトライフォン“s;a;h”と、図４に示す言語モデルの情報から得られた単語「の(n;o)」とその前に続く単語「朝(a;s;a)」との連鎖「朝の(a;s;a;n;o)」における３番目の音

素/ａ/とその前後に続く音素とから成るトライフォン“s;a;n”とについて、音素仮説を展開することが可能である。この場合には２つの音素仮説を展開するだけでよいが、より複雑な文法や統計的言語モデルを参照した場合には単語の終端で多くの次の単語につながる可能性があり、図５(b)に示すように、次の単語の先頭音素に応じて多数の音素仮説を展開することになる。これに対して、本実施の形態のように音素状態木の音素仮説を展開する場合には、次の単語の先頭音素に関係なく図２(a)に示すような音素状態木“s;a;＊”を、図５(a)に示すように１つ展開するだけでよいのである。尚、図５(a)においては、音素状態木のシンボルとして「木」を模した三角形を当てている。

【００３７】

ところで、図５(b)に示すように、個々の音素について仮説を展開する場合には、次に続く単語の先頭音素の種類を全２７とした場合、新たに展開される音素仮説の数は２７となり、全音素仮説における状態の総数は８１(＝２７×３)となる。

【００３８】

これに対して、図５(a)に示すように、上記音素状態木を用いて音素仮説を展開することによって、新たに展開される音素仮説の数は１となり、状態の総数は２９(１＋７＋２１)に削減することができる。したがって、仮説の展開処理および照合処理の処理量を大幅に削減できるのである。

【００３９】

また、上記言語モデルに文法を用いる場合、単語辞書４および言語モデルによって後続の音素が限定されることが多い。そこで、図８に示すように、音素状態木“s;a;＊”の各状態のうち、単語辞書４に基づく音素列“s;a;h”および言語モデルに基づく音素列“s;a;n”に必要な状態のみにフラグ(図８中においては楕円印)を付すことによって、照合の全状態数を、音素状態木“s;a;＊”の総ての状態数２９に比して状態数５に削減できる。したがって、照合の処理量を更に削減できるのである。

【００４０】

以上のごとく、本実施の形態においては、音素環境依存音響モデル格納部３に

は、先行音素および中心音素が同じトライフォンモデルの状態系列をまとめて木構造化した音素状態木を格納している。その結果、複数のトライフォンモデルで状態を共有している状態共有モデルの場合には、木構造化した際に共有されている状態を一つにまとめることができ、ノード数を削減することができる。したがって、個々の音素について仮説を展開する場合に上記音素状態木を音素仮説として用いることによって、次に続く単語の先頭音素に関係無く１つの音素仮説を展開すればよいことになる。したがって、次に続く単語の先頭音素の種類を全２７と仮定した場合、従来は、新たに２７個の音素仮説が展開されるために全音素仮説における状態の総数は８１となる。これに対して、本実施の形態においては、新たに展開される音素仮説は１個であるために全音素仮説における状態の総数を２９に削減することができるのである。

【００４１】

すなわち、本実施の形態によれば、上記前向き照合部２によって、音素環境依存音響モデル格納部３に格納された音素環境依存音響モデル，言語モデル格納部５に格納された言語モデルおよび単語辞書４を参照して音素仮説を展開する際における音素仮説の展開処理量を大幅に削減できる。したがって、単語内および単語境界に関係なく、仮説の展開が容易になる。また、前向き照合部２によって、上記音素環境依存音響モデルを用いて、音響分析部１からの特徴パラメータ系列と上記展開された音素仮説とのフレーム同期ビタビビームサーチによる照合を行う際における照合処理量を大幅に削減できるのである。

【００４２】

また、その際に、上記前向き照合部２は、上記音素仮説との照合を行う際に、各音素仮説のスコアを計算し、スコアの閾値あるいは仮説数の閾値に基づいて音素仮説の枝刈りを行うようにしている。したがって、単語となる可能性が低い音素仮説を削除することができ、照合処理量を大幅に削減することができる。さらに、前向き照合部２は、上記音素仮説を展開する際に、言語モデル格納部５および単語辞書４を参照して、上記音素仮説を構成する音素状態木の状態のうち、互いに接続可能であって上記照合に関係のある状態のみにフラグを付けるようにすることができる。したがって、その場合には、木構造化された状態のうち上記照

合に関係のない状態に関するビタビ計算を行う必要がなく、照合処理量を更に削減することができるのである。

【００４３】

尚、上述の説明において、上記音素環境依存音響モデルは、トライフォンモデルと呼ばれる前後１つずつの音素環境を考慮したＨＭＭを用いたが、隣接するサブワードに依存して決定されるサブワードはこれに限定されるものではない。

【００４４】

ところで、上記実施の形態における音響分析部１,前向き照合部２および後向き探索部８による上記音響分析手段,照合手段および検索手段としての機能は、プログラム記録媒体に記録された連続音声認識プログラムによって実現される。上記実施の形態における上記プログラム記録媒体は、ＲＡＭ(ランダム･アクセス･メモリ)とは別体に設けられたＲＯＭ(リード･オンリ･メモリ)でなるプログラムメディアである。あるいは、外部補助記憶装置に装着されて読み出されるプログラムメディアであってもよい。尚、何れの場合においても、上記プログラムメディアから連続音声認識プログラムを読み出すプログラム読み出し手段は、上記プログラムメディアに直接アクセスして読み出す構成を有していてもよいし、上記ＲＡＭに設けられたプログラム記憶エリア(図示せず)にダウンロードし、上記プログラム記憶エリアにアクセスして読み出す構成を有していてもよい。尚、上記プログラムメディアからＲＡＭの上記プログラム記憶エリアにダウンロードするためのダウンロードプログラムは、予め本体装置に格納されているものとする。

【００４５】

ここで、上記プログラムメディアとは、本体側と分離可能に構成され、磁気テープやカセットテープ等のテープ系、フロッピーディスク,ハードディスク等の磁気ディスクやＣＤ(コンパクトディスク)－ＲＯＭ,ＭＯ(光磁気)ディスク,ＭＤ(ミニディスク),ＤＶＤ(ディジタル多用途ディスク)等の光ディスクのディスク系、ＩＣ(集積回路)カードや光カード等のカード系、マスクＲＯＭ,ＥＰＲＯＭ(紫外線消去型ＲＯＭ),ＥＥＰＲＯＭ(電気的消去型ＲＯＭ),フラッシュＲＯＭ等の半導体メモリ系を含めた、固定的にプログラムを坦持する媒体である。

【００４６】

また、上記実施の形態における連続音声認識装置は、モデムを備えてインターネットを含む通信ネットワークと接続可能な構成を有する場合には、上記プログラムメディアは、通信ネットワークからのダウンロード等によって流動的にプログラムを担持する媒体であっても差し支えない。尚、その場合における上記通信ネットワークからダウンロードするためのダウンロードプログラムは、予め本体装置に格納されているものとする。あるいは、別の記録媒体からインストールされるものとする。

【００４７】

尚、上記記録媒体に記録されるものはプログラムのみに限定されるものではなく、データも記録することが可能である。

【００４８】

【発明の効果】

以上より明らかなように、第１の発明の連続音声認識装置は、照合部で、環境依存音響モデルの状態系列のうち、複数のサブワードモデルの状態系列をまとめて木構造化して成るサブワード状態木,単語辞書および言語モデルを参照してサブワードの仮説を展開すると共に、音響分析部からの特徴パラメータの時系列と上記展開された仮説との照合を行って、単語の終端に該当する仮説に関する単語,累積スコアおよび始端開始フレームを含む単語ラティスを出力するので、次に続く単語の先頭サブワードに関係無く１つの仮説を展開すればよく、全仮説における状態の総数を削減することができる。

【００４９】

したがって、上記仮説の展開処理量を大幅に削減でき、単語内および単語境界に関係なく、上記仮説の展開を容易に行うことができる。さらに、上記照合を行う際における照合処理量を大幅に削減することができる。

【００５０】

また、１実施例の連続音声認識装置は、上記環境依存音響モデルを、先行サブワードおよび中心サブワードが同じサブワードモデルの状態系列を木構造化したサブワード状態木としたので、次の仮説を展開する場合には、終端仮説における中心サブワードのみに注目して対応する先行サブワードを有するサブワード状態

木を展開すればよい。したがって、後続サブワードが複数あってもより少ない仮説を展開すればよく、仮説の展開を容易にできる。

【００５１】

また、１実施例の連続音声認識装置は、複数のサブワードモデルで状態を共有している状態共有モデルを木構造化したサブワード状態木を環境依存音響モデルとしたので、後段のサブワードによって共有される前段のサブワードの状態を一つにまとめてノード数を削減することができる。したがって、上記照合時における処理量を大幅に削減できる。

【００５２】

また、１実施例の連続音声認識装置は、上記照合部を、上記仮説の展開を行う際に、上記単語辞書および言語モデルから得られる接続可能なサブワード情報を用いて、上記仮説であるサブワード状態木を構成する状態のうち、互いに接続可能な状態にフラグを付すので、上記照合の際にビタビ計算を行う必要がある状態を限定して、照合処理量を更に簡単にできる。

【００５３】

また、１実施例の連続音声認識装置は、上記照合部を、上記照合を行う際に、上記特徴パラメータの時系列に基づいて算出された上記仮説のスコアの閾値あるいは仮説数を含む基準に従って、上記仮説の枝刈りを行うようにしたので、単語となる可能性が低い仮説を削除して、以後の照合処理量を大幅に削減できる。

【００５４】

また、第２の発明の連続音声認識方法は、音素環境依存音響モデルの状態系列のうち、複数のサブワードモデルの状態系列をまとめて木構造化して成るサブワード状態木,単語辞書および言語モデルを参照してサブワードの仮説を展開すると共に、特徴パラメータの時系列と上記展開された仮説との照合を行って、単語の終端に該当する仮説に関する単語,累積スコアおよび始端開始フレームを含む単語ラティスを出力するので、上記第１の発明の場合と同様に、次に続く単語の先頭サブワードに関係無く１つの仮説を展開すればよく、全仮説における状態の総数を削減することができる。

【００５５】

したがって、上記仮説の展開処理量を大幅に削減でき、単語内および単語境界に関係なく、上記仮説の展開を容易に行うことができる。さらに、上記照合を行う際における照合処理量を大幅に削減することができる。

【００５６】

また、第３の発明の連続音声認識プログラムは、コンピュータを、上記第１の発明における音響分析部,単語辞書,言語モデル格納部,環境依存音響モデル格納部,照合部及び探索部として機能させるので、上記第１の発明の場合と同様に、次に続く単語の先頭サブワードに関係無く１つの仮説を展開すればよく、単語内および単語境界に関係なく仮説の展開を容易にできる。さらに、特徴パラメータ系列と上記展開された仮説との照合を行う際における照合処理量を大幅に削減できる。

【００５７】

また、第４の発明のプログラム記録媒体は、上記第３の発明の連続音声認識プログラムが記録されているので、上記第１の発明の場合と同様に、次に続く単語の先頭サブワードに関係無く１つの仮説を展開すればよく、単語内および単語境界に関係なく仮説の展開を容易にできる。さらに、特徴パラメータ系列と上記展開された仮説との照合を行う際における照合処理量を大幅に削減できる。

【図面の簡単な説明】

【図１】　この発明の連続音声認識装置におけるブロック図である。

【図２】　音素環境依存音響モデルの説明図である。

【図３】　図１における単語辞書の説明図である。

【図４】　言語モデルの説明図である。

【図５】　図１における前向き照合部による仮説の展開の説明図である。

【図６】　上記前向き照合部によって実行される前向き照合処理動作のフローチャートである。

【図７】　上記前向き照合部による仮説の照合および仮説の枝刈りの説明図である。

【図８】　音素仮説の音素状態木における必要な状態のみにフラグを付す場合の説明図である。

【図９】　認識単語と単語間単語との境界の履歴が考慮されない場合と考慮された場合との比較図である。

【符号の説明】

１…音響分析部、

２…前向き照合部、

３…音素環境依存音響モデル格納部、

４…単語辞書、

５…言語モデル格納部、

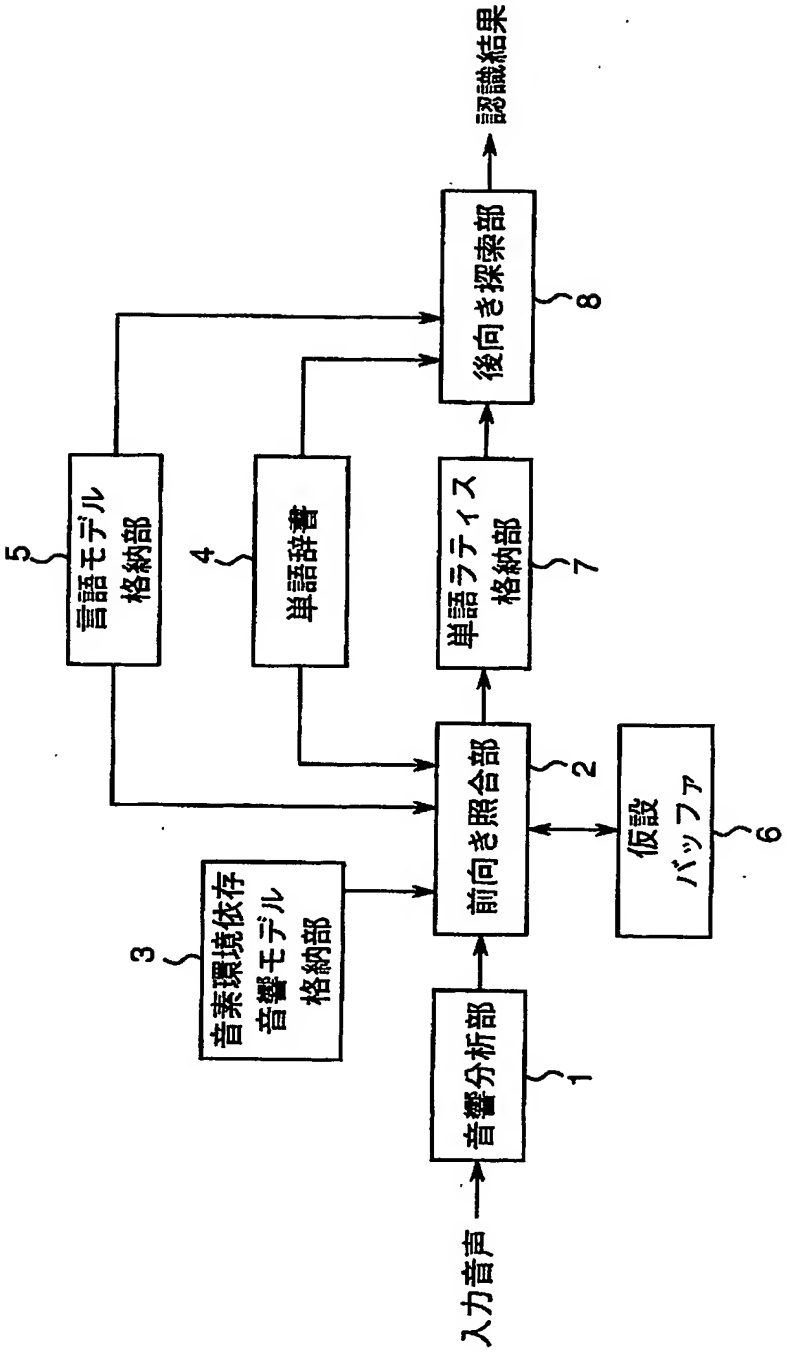
６…仮説バッファ、

７…単語ラティス格納部、

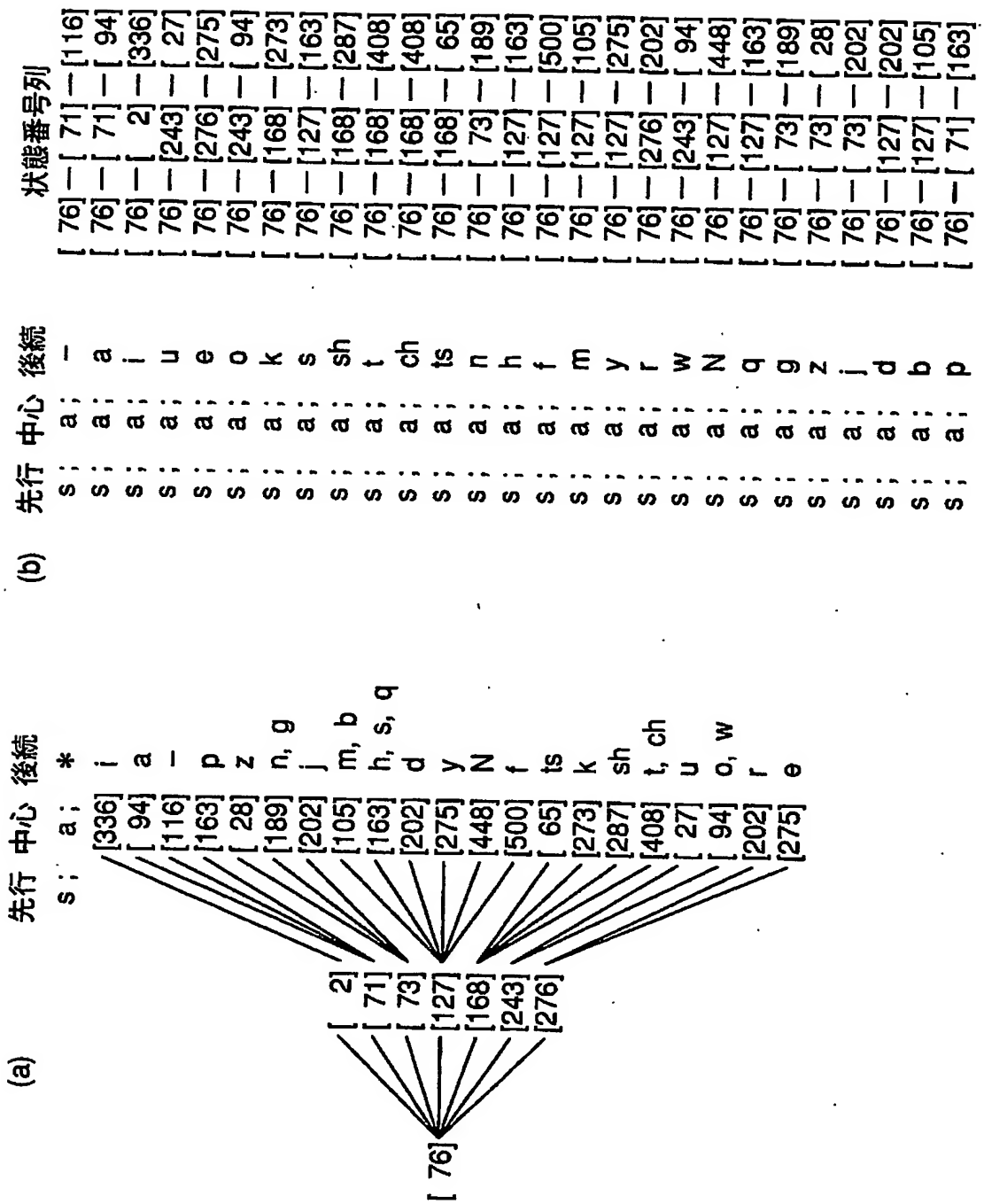
８…後向き探索部。

【書類名】　図面

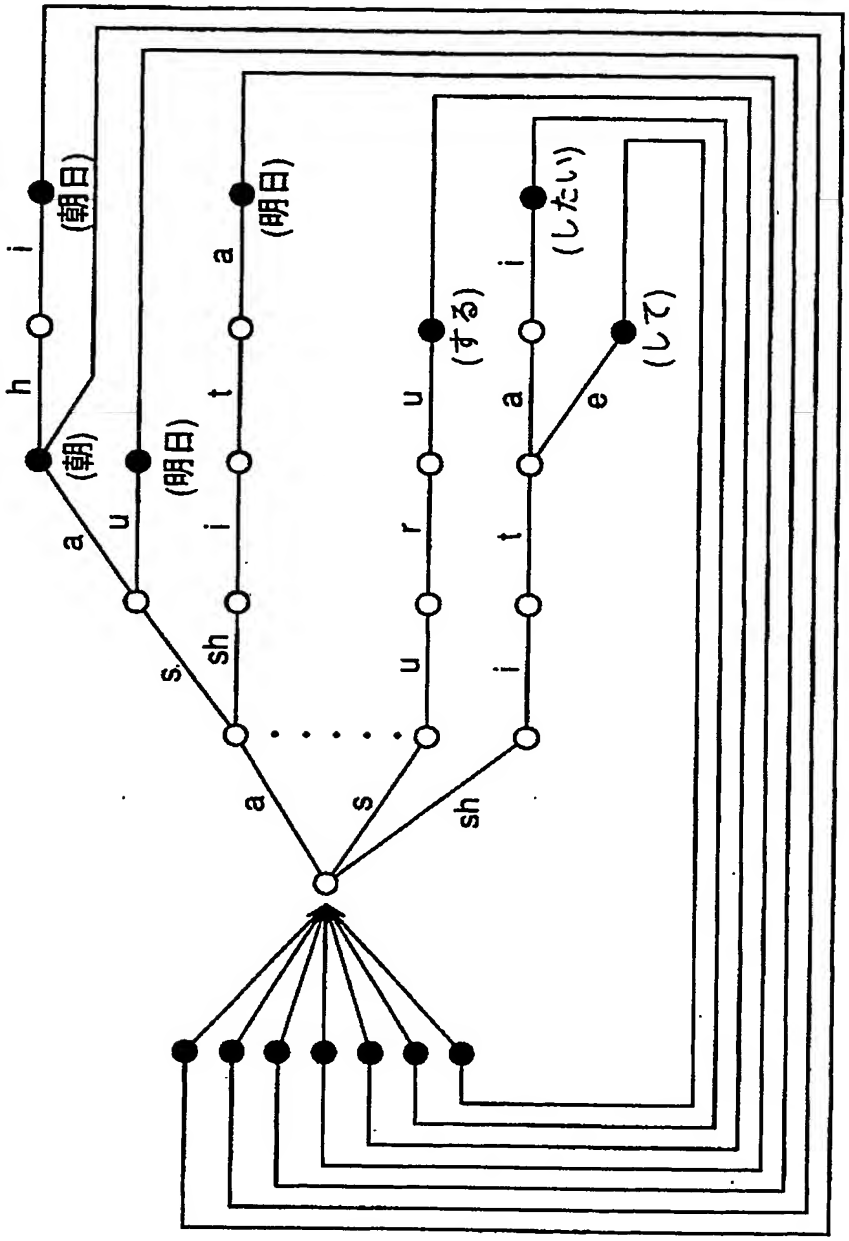
【図１】


入力音声
音響分析部
1
前向き照合部
2
音素環境依存
音響モデル
格納部
3
単語辞書
4
音語モデル
格納部
5
仮設
バッファ
6
単語ラティス
格納部
7
後向き探索部
8
認識結果

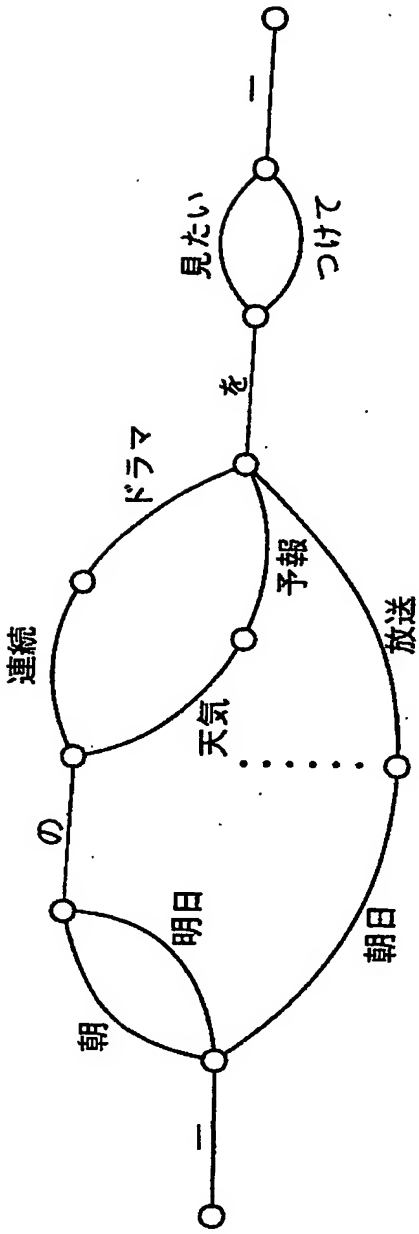
【図２】


(a)
先行 中心 後続
s ; a ; *
[336] i
[94] a
[116] –
[163] p
[28] z
[189] n, g
[202] j
[105] m, b
[163] h, s, q
[202] d
[275] y
[448] N
[500] f
[65] ts
[273] k
[287] sh
[408] t, ch
[27] u
[94] o, w
[202] r
[275] e
[2]
[71]
[73]
[127]
[168]
[243]
[276]
[76]
(b)
先行 中心 後続
状態番号列
s ; a ; – [76] — [71] — [116]
s ; a ; a [76] — [71] — [94]
s ; a ; i [76] — [2] — [336]
s ; a ; u [76] — [243] — [27]
s ; a ; e [76] — [276] — [275]
s ; a ; o [76] — [243] — [94]
s ; a ; k [76] — [168] — [273]
s ; a ; s [76] — [127] — [163]
s ; a ; sh [76] — [168] — [287]
s ; a ; t [76] — [168] — [408]
s ; a ; ch [76] — [168] — [408]
s ; a ; ts [76] — [168] — [65]
s ; a ; n [76] — [73] — [189]
s ; a ; h [76] — [127] — [163]
s ; a ; f [76] — [127] — [500]
s ; a ; m [76] — [127] — [105]
s ; a ; y [76] — [127] — [275]
s ; a ; r [76] — [276] — [202]
s ; a ; w [76] — [243] — [94]
s ; a ; N [76] — [127] — [448]
s ; a ; q [76] — [127] — [163]
s ; a ; g [76] — [73] — [189]
s ; a ; z [76] — [73] — [28]
s ; a ; j [76] — [73] — [202]
s ; a ; d [76] — [127] — [202]
s ; a ; b [76] — [127] — [105]
s ; a ; p [76] — [71] — [163]

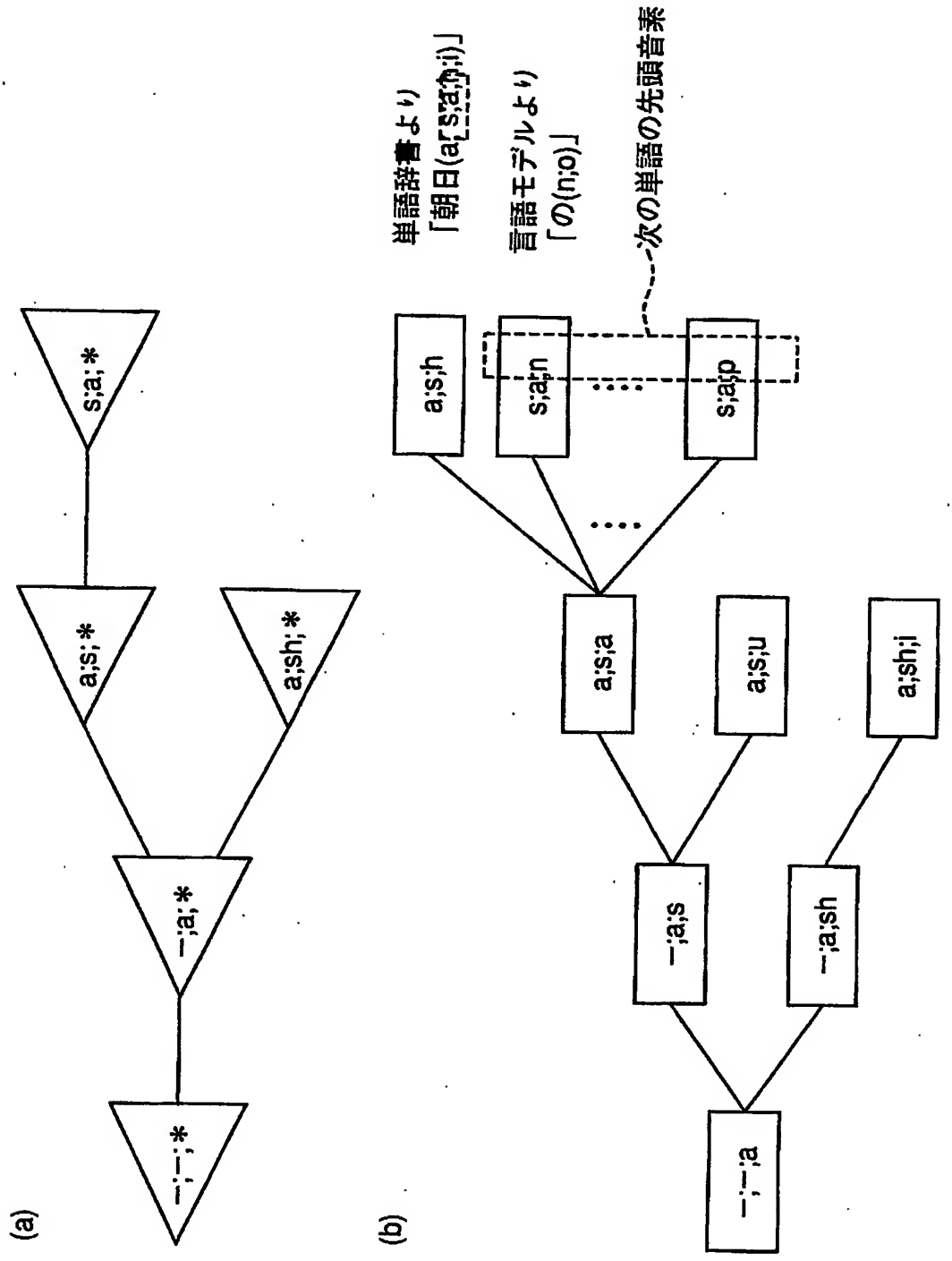
【図３】



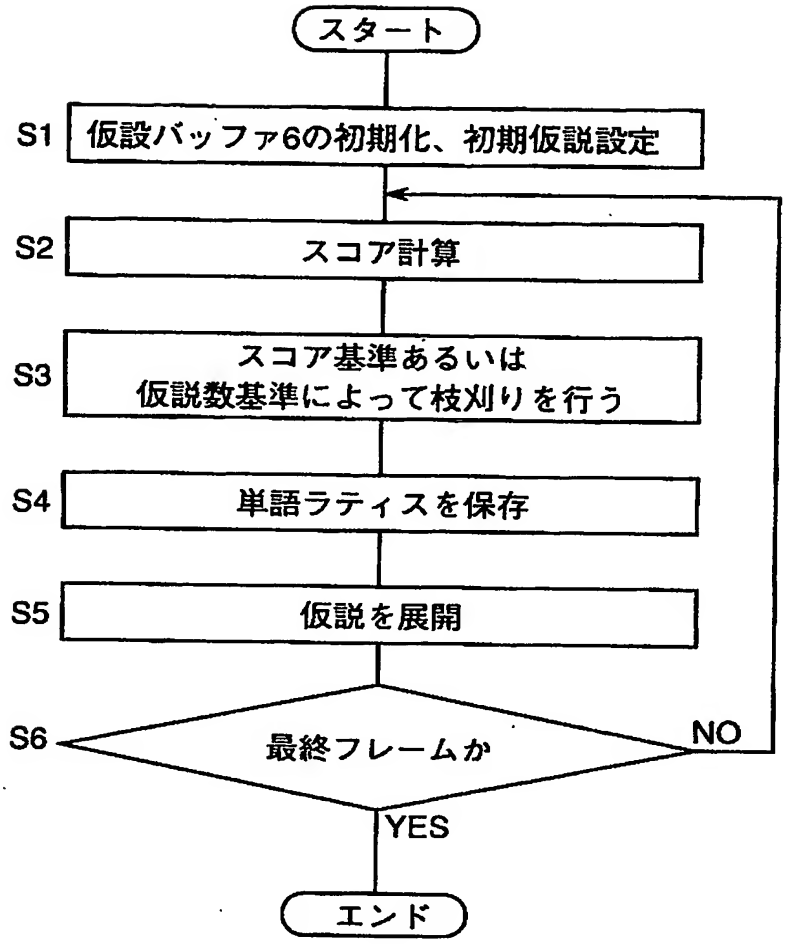
【図４】



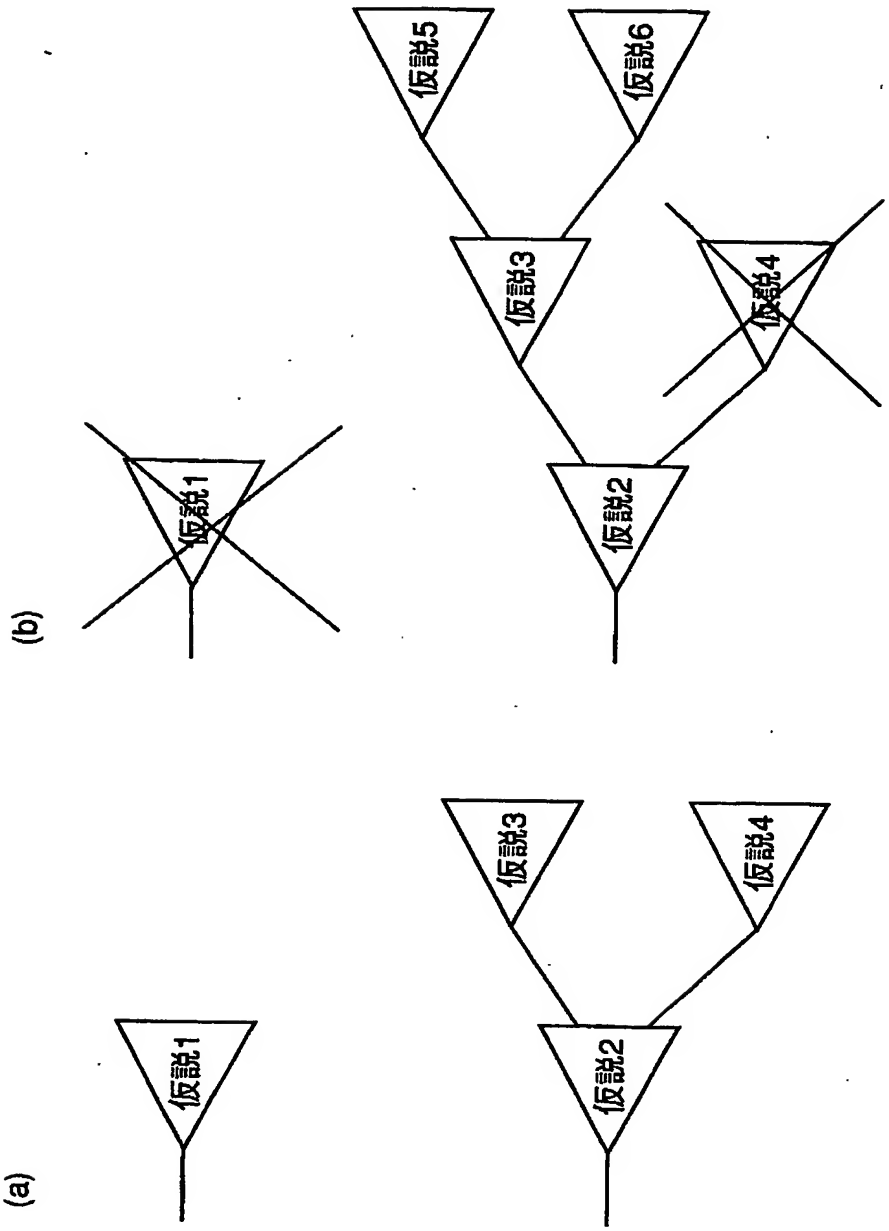
【図５】


(a)
s;a;*
a;s;*
a;sh;*
−;a;*
−;−;*
(b)
単語辞書より
「朝日(a;s;a;sh;i)」
言語モデルより
「の(n;o)」
次の単語の先頭音素
a;s;h
s;a;n
s;a;p
a;s;a
a;s;u
a;sh;i
−;a;s
−;a;sh
−;−;a

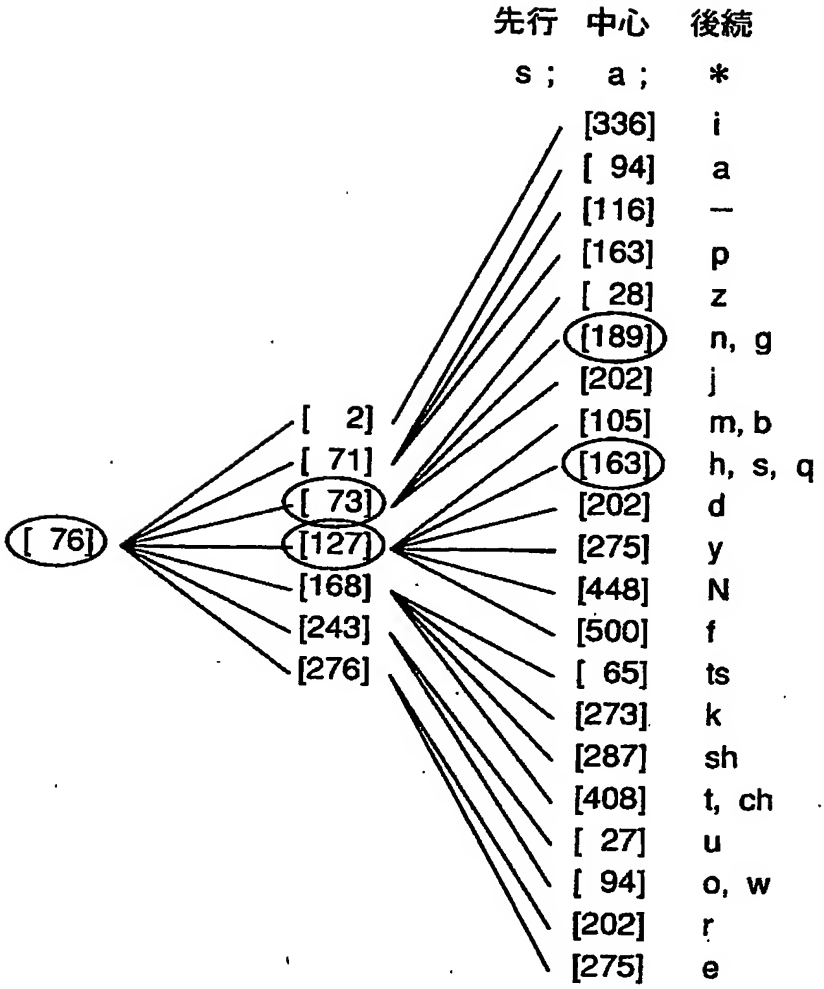
【図６】


スタート
S1
仮設バッファ6の初期化、初期仮説設定
S2
スコア計算
S3
スコア基準あるいは
仮説数基準によって枝刈りを行う
S4
単語ラティスを保存
S5
仮説を展開
S6
最終フレームか
NO
YES
エンド

【図７】

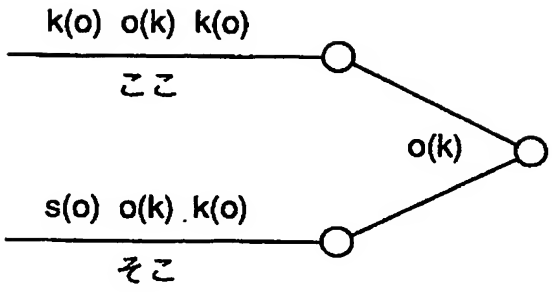


【図８】

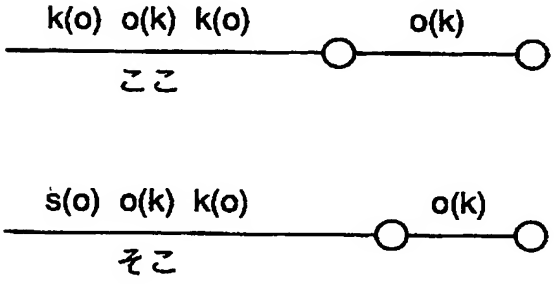

先行
s ;
中心
a ;
後続
*
[336] i
[94] a
[116] —
[163] p
[28] z
[189] n, g
[202] j
[105] m, b
[163] h, s, q
[202] d
[275] y
[448] N
[500] f
[65] ts
[273] k
[287] sh
[408] t, ch
[27] u
[94] o, w
[202] r
[275] e
[2]
[71]
[73]
[127]
[168]
[243]
[276]
[76]

【図９】

(a)



(b)



【書類名】　要約書

【要約】

【課題】　単語境界にも音素環境依存音響モデルを用いて精度を確保しつつ大語彙の連続音声認識時にも処理量の増大を抑える。

【解決手段】　音素環境依存音響モデル格納部３には、先行音素および中心音素が同じトライフォンモデルをまとめて先行音素の状態と中心音素の状態と後続音素の状態との状態系列を木構造化した音素状態木を格納している。したがって、前向き照合部２によって、上記音素状態木,言語モデル格納部５に格納された言語モデルおよび単語辞書４を参照して音素仮説を展開する際には、次に続く単語の先頭音素に関係無く１つの音素仮説を展開すればよく、単語内および単語境界に関係なく仮説の展開が容易になる。また、音響分析部１からの特徴パラメータ系列との照合を行う際における照合処理量を大幅に削減できる。

【選択図】　図１

出 願 人 履 歴 情 報

識別番号　〔０００００５０４９〕

１．変更年月日　１９９０年　８月２９日
［変更理由］　新規登録
住　所　大阪府大阪市阿倍野区長池町２２番２２号
氏　名　シャープ株式会社